APPENDIX A

USE OF INDIVIDUAL SAMPLES IN FISH CONTAMINANT MONITORING PROGRAMS

APPENDIX A

USE OF INDIVIDUAL SAMPLES IN FISH CONTAMINANT MONITORING PROGRAMS

The use of composite samples is often the most cost-effective method for estimating average tissue concentrations of analytes in target species populations to assess chronic human health risks. However, there are some situations in which individual sampling can be more appropriate from both ecological and risk assessment perspectives. Individual sampling provides a direct measure of the range and variability of contaminant levels in target fish populations. Information on maximum contaminant concentrations in individual fish is useful in evaluating acute human health risks. Estimates of the variability of contaminant levels among individual fish can be used to ensure that studies meet desired statistical objectives. For example, the population variance of a contaminant can be used to estimate the sample size needed to detect statistically significant differences in the mean contaminant concentration compared to the contaminant screening values. Finally, the analysis of individual samples may be desirable, or necessary, when the objective is to minimize the impacts of sampling on certain vulnerable target populations, such as predators in headwater streams and aquatic turtles, and in cases where the cost of collecting enough individuals for a composite sample is excessive.

Analyzing individual fish incurs additional expenses, particularly when one considers that a number of individual analyses are required to achieve measurements of a reasonable statistical power. However, the recommendation that States archive the individual fish homogenates from which composite samples are prepared for both screening and intensive studies (see Section 6.1.1.6) would make it possible to perform individual analyses where needed without incurring additional sampling costs.

Individual analysis is especially well-suited for intensive studies, in which results from multiple stations and time periods are to be compared. The remainder of this appendix discusses how the sampling design might be affected by analyzing individual rather than composite samples and how contaminant data from individuals versus composites might be used in risk assessments.

A.1 SAMPLING DESIGN

There are seven major components of the sampling design for a fish or shellfish monitoring program: site selection, target species, target analytes, target analyte screening values (SVs), sampling time, sampling type and size class, and replicate samples. Of these, only the number of replicate samples and possibly the

target species would be expected to differ if individual samples were analyzed rather than composites. Target species becomes a limiting factor when individuals of the target species are not large enough to provide adequate tissue mass for all the required chemical analyses.

The five factors that determine the optimal number of fish or shellfish to analyze are presented in Section 6.1.2.7. Briefly, the five factors are:

- Cost components
- Minimum detectable difference between site-specific mean target analyte concentration and SV
- Level of significance
- Population variance
- Power of the hypothesis test

Each of these characteristics will be examined in detail for the collection and analysis of individual samples.

A.1.1 Cost Components

The cost of obtaining contaminant data from individual fish or shellfish is compared to the cost of obtaining contaminant data from composite samples in Table A-1. These costs are dependent on the separate costs of collecting, preparing, and analyzing the samples.

Typically, the cost of collecting individual samples will be less than that of collecting composite samples when the target species is scarce or difficult to capture. The cost of collecting individuals may not be a factor if the sample

Table A-1 Relative Cost of Obtaining Contaminant Data from Individual Versus Composite Samples

	Relative cost	
Cost component	**Composite samples**	**Individual samples**
Collection	Moderate to high	Low to moderate
Preparation	Very low to moderate	Very low to low
Analysis	Low to moderate	Moderate to high

collection method used typically allows for the collection of a large number of individuals in a short period of time. In some situations, seines or gill nets might have this characteristic. Also, in estuaries, coastal water, or large lakes where productivity is high, the additional cost of collecting large numbers of individuals for composite sampling may be minimal compared to the effort expended for collecting individual samples.

The cost of preparing individual samples for analysis is typically lower than either the costs of collection or analysis. Generally, the cost of preparing composite samples for analysis will be greater than that of preparing individual samples. Sample preparation procedures can range in complexity from the grinding of whole fish to delicate and time-consuming operations to resect specific tissues. Costs of composite sampling depend largely on the number of individuals required per composite sample and the number of replicate composite samples required to achieve the desired statistical power; however, these costs can be somewhat controlled (see Section 6.1.2.7).

The cost of analyzing individual samples is also typically higher than the cost of analyzing composite samples. The cost differential between the two approaches is directly correlated to the cost for the analysis of a single sample. For some intensive studies, the number of target analytes exceeding the SV is small, so few analyses are required. In these cases, the relative costs between the two approaches may not differ greatly if the number of samples analyzed using the two different approaches is similar (e.g., three to five samples). A sampling design with such a small number of individual samples would be appropriate only if the expected mean target analyte concentration was much greater than the SV.

A.1.2 Minimum Detectable Difference

The difference between the mean target analyte concentration at a site and the SV will not often be known before the screening study has been performed. The minimum detectable difference between the mean concentration and the SV will depend on the level of significance (see Section A.1.3), population variance (Section A.1.4), and the number of replicates collected. In practice, the sample size is often determined by establishing the minimum detectable difference prior to the study according to the objectives of the project. For an SV that has not been multiplied by an uncertainty factor, the cost of detecting a 10 percent difference may be warranted. The issue of minimum detectable difference is discussed in greater detail in Section A.1.5.

A.1.3 Level of Significance

The level of significance (LS) refers to the probability of incorrectly rejecting the null hypothesis, that there is no difference between the mean target analyte concentration and the SV. This probability is also called Type I error. The LS can be thought of as the chance of a "false positive" or of detecting a difference that does not exist. The LS affects the sampling design by modifying the required

power (thus impacting the sample size) of the statistical test to detect a significant difference between the mean target analyte concentration and the SV (see Section A.1.5). A typical LS used in biological sampling is 0.05. In some cases, an LS other than 0.05 could be appropriate. If the ramifications of a statistically significant difference are severe, a more conservative LS (e.g. 0.01) might be used. On the other hand, if the statistical test is being conducted to identify whether additional sampling should be performed (i.e., a screening survey), then a less conservative LS (e.g. 0.10) might be used.

A.1.4 Population Variance

The variability in target analyte concentrations within a given fish or shellfish population is a critical factor in determining how many individual samples to collect and analyze. The population variance directly affects the power of the statistical test to detect a significant difference between the mean target analyte concentration and the SV (see Section A.1.5) by impacting the sample size. The population variance may not be known prior to sampling, but it can be estimated from similar data sets from the same target species, which could in many cases be obtained by analyzing individual fish homogenates if these have been archived as recommended in Section 6.1.1.6. In using historical data to estimate population variance, it is important to consider contaminant data only from individual fish or shellfish of the same species. By its very nature, a data set consisting of replicate composite samples tends to smooth out the variability inherent in a group of individual organisms. An extreme example of this phenomenon was presented by Fabrizio et al. (1995) in a study on procedures for compositing fish samples. They used computer simulations to predict PCB concentrations in composite samples of striped bass that had previously been analyzed individually. The predicted variance in these concentrations in the composite samples was approximately 20 percent of the variance obtained from individual analyses.

A.1.5 Power of Statistical Test

Another critical factor in determining the sample size is the power of the statistical test, that is, the probability of detecting a true difference between the mean target analyte concentration and the SV. Because of its profound influence on sample size, it is the power of the test that may ultimately control whether the objectives of the survey are met. The effect of joint consideration of the desired power, the population variance, and the minimum detectable difference on the sample size is described by the following formula (Steel and Torrie, 1980):

$$n = \frac{(Z_\alpha + Z_\beta)^2 \, 2\sigma^2}{\delta^2}$$

where

n = sample size
Z_α = Z statistic for Type I error (α)
Z_β = Z statistic for Type II error (β)
σ^2 = population variance (estimated from historical data)
δ = minimum detectable difference between mean target analyte concentration and SV.

Recall that the Type I error is equal to the LS, and the value is generally between 0.01 and 0.10. Type II error is the probability of accepting the null hypothesis (that there is no difference between the mean target population concentration and the SV) when it is actually false. This type of error can be thought of as the chance of a "false negative," or not detecting a difference that does in fact exist. The complement of Type II error (1-β) is the power of the statistical test.

The above equation for determining sample size was solved for powers ranging from 0.5 to 0.9 (50 to 90 percent; Figure A-1) assuming an LS of 0.05. The values for σ (standard deviation) and δ were set relative to the SV. A similar exercise was performed in Section 6.1.2.7 and two examples were provided. In example A, both the standard deviation and minimum detectable difference were set to 0.5 SV. Example A corresponds to a ratio of 1 on the x-axis of Figure A-1. Applying example A to the collection of individual fish, the recommended sample size would range from approximately 6 individual samples for a power of 50 percent to 18 individual samples for a power of 90 percent (Figure A-1). In example B, the standard deviation was set to 1.0 SV, while the minimum detectable difference was kept at 0.5 SV. Example B corresponds to a ratio of 2 on the x-axis of Figure A-1. Applying example B to the collection of individual samples, the sample size would have to be almost 40 individual samples to achieve even a modest statistical power (i.e., 70 percent).

It is common to set the power of the statistical test to at least 80 percent (Fairweather, 1991). Figure A-1 indicates that, to achieve a statistical power of 80 percent using the variability assumptions in examples A and B, 13 and 50 fish would have to be collected, respectively. The estimated sample sizes for individual fish or shellfish is similar to those calculated for composite samples (see Section 6.1.2.7). For example A as applied to composite samples, 12 to 18 fish would have to be collected. For example B as applied to composite samples, 30 to 50 fish would have to be collected. Thus, the cost of collecting the fish to achieve a power of 80 percent would not be significantly different for composite versus individual samples (see Section A.1.1). The number of

Figure A-1. Recommended sample sizes to achieve various statistical powers.

analyses, however, would be considerably less for composite samples (3 to 10 analyses of composite samples versus 13 or 50 analyses of individual samples).

Figure A-1 also indicates that 10 or fewer individual fish or shellfish should be analyzed only if the ratio of the standard deviation to the minimum detectable difference is 0.85 or less. For ratios less than 0.5, the effect of sample size on the statistical power is minor. If the expected mean target analyte concentration is many times greater than the SV, it may not be necessary to allocate resources toward the collection and analysis of more than a minimum number (e.g., three to five samples) of individual fish or shellfish.

A.2 USE OF CONTAMINANT DATA FROM INDIVIDUAL FISH/SHELLFISH IN RISK ASSESSMENTS

Target analyte concentrations in composite samples represent averages for specific target species populations. The use of these values in risk assessments is appropriate if the objective is to estimate the average concentration to which consumers of the target species might be exposed over a long period of time. The use of long exposure durations (e.g., 30 to 70 years) is typical of the assessment of carcinogenic target analytes, the health effects of which may be manifested over an entire lifetime (see Volume II of this series). Target analytes that produce noncarcinogenic effects, on the other hand, may cause acute effects to human health over a relatively short period of time on the order of hours or days. The use of average contaminant concentrations derived from the analysis of composite samples may not be protective against acute health effects because high concentrations in an individual organism may be masked by lower concentrations in other individuals in the composite sample. Contaminant data from individual samples permits the use of alternative estimates of contaminant concentration for a group of fish or shellfish (e.g., maximum). Therefore, the decision whether to collect and analyze individual fish or shellfish may depend on the target analytes included in the monitoring program.

EPA has recommended that 25 target analytes be included in screening studies (see Section 4). All of the target analytes except PCBs, PAHs, and dioxins/furans have reference doses for noncarcinogenic health effects, although the carcinogenic risk is likely to be greater than the noncarcinogenic risk for eight other target analytes (see Table 5-2). EPA's draft reassessment of the health effects of 2,3,7,8-TCDD (dioxin) indicated that this chemical may also pose a significant noncarcinogenic health risk in some cases (U.S. EPA, 1994).

A.3 EXAMPLE CASE STUDY

The presentation of a case study will illustrate some of the sample size and data interpretation issues discussed in Sections A.1 and A.2, respectively. A State has prepared a composite sample of target species A from a particular water-body of concern. This composite sample was analyzed for all 25 target analytes listed in Table 4-1. Of the 25 target analytes, only cadmium was detected at a concentration exceeding the SV (10 ppm) for cadmium listed in Table 5-2.

Cadmium was detected at 20 ppm, twice the SV calculated for cadmium. Because the SV for at least one target analyte was exceeded, an intensive study was warranted. The State decided to collect and analyze individual fish in the intensive study for the following reasons: (1) the cost of collecting individual fish is less than the cost of collecting fish for composites, (2) the analytical costs for analyzing cadmium are relatively low (<$50 sample), and (3) the cadmium concentrations in individual fish should more accurately reflect the potential acute (noncarcinogenic) health risk from cadmium than the mean cadmium concentration derived from composite samples.

The first issue the State must decide is how many individual fish to collect and analyze. The important factors in this decision are the minimum detectable difference the State wishes to test and the variability in cadmium concentrations within the target species population. The first factor can be obtained from the results of the screening survey. The State wishes to test whether the difference between the concentration detected in the single composite sample (20 ppm) and the SV (10 ppm) is significant. This assumes that the mean cadmium concentration for the individual is also 20 ppm. The expected standard deviation (8 ppm) was obtained from a previous investigation performed on individuals of the target species and was equal to 0.8 of the SV (10 ppm). Using Figure A-1, it can be seen that, for a ratio of standard deviation (0.8 x SV) to detectable difference (1.0 x SV) of 0.8, the sample size necessary to achieve a statistical power of 80 percent would be eight fish.

The State determines that the mean cadmium concentration of eight individual fish of the target species is 30 ppm and the standard deviation is equal to the predicted value of 8 ppm. The State performs a *t*-test to determine if the mean concentration is significantly greater than the SV. As described in Section 6.1.2.7, the statistic

(mean - SV)/standard deviation

has a *t*-distribution with n-1 degrees of freedom. For this example, the *t* statistic is 2.5 ([(30-10)/8] with 7 degrees of freedom. This value exceeds the critical t-statistic (1.895) for a one-tailed LS of 0.05. Therefore, the State determines that the mean cadmium concentration for these eight individual fish of the target species is significantly greater than the SV and a risk assessment is performed.

A.4 REFERENCES

Fabrizio, M.C., A.M. Frank, and J.F. Savino. 1995. Procedures for formation of composite samples from segmented populations. *Environmental Science and Technology* 29(5):1137-1144.

Fairweather, P.G. 1991. Statistical power and design requirements for environmental monitoring. *Aust. J. Freshwater Res.* 42:555-567.

Steel, R.G.D., and J.H. Torrie. 1980. *Principles and Procedures of Statistics. A Biometrical Approach.* Second Edition. McGraw-Hill Book Company. New York, NY. 633 pp.

U.S. EPA (U.S. Environmental Protection Agency). 1989. *Risk Assessment Guidance for Superfund: Volume I, Human Health Evaluation Manual.* Office of Emergency and Remedial Response, Washington, DC. EPA/540/1-89/002.

U.S. EPA (U.S. Environmental Protection Agency). 1994. Health Assessment for 2,3,7,8-TCDD and Related Compounds. Public Review Draft. EPA/600/EP-92/001.